DOCUMENT RESUME

ED 206 714 TM 810 619

AUTHOR Logsdon, David M.
TITLE An Experimental Comparison of the Effectiveness of Two Domain Definition Strategies for Criterion-Referenced Tests.
PUB DATE 81
NOTE 15p.: Paper presented at the Annual Meeting of the Eastern Educational Research Association (Philadelphia, PA, March, 1981).

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Comparative Analysis; *Criterion Referenced Tests; Methods; *Test Construction; *Test Items
IDENTIFIERS *Domain Specifications; *Instructional Objectives Exchange; Item Homogeneity

ABSTRACT

This study examined the effectiveness of two different domain definition strategies in achieving homogeneity of criterion-referenced test items. The argument was tested regarding the extent to which item writers following the Instructional Objectives Exchange (IOX) domain definition strategy for a cognitive skill generate items that are more homogeneous than those generated from the simple objectives strategy for the same skill. An issue of concern was that of the quantity and quality of specifications needed to achieve clarity of definition and precision among the items. The IOX strategy contains four parts: general description, sample item, stimulus attributes, and response attributes. Results showed the IOX specifications as laborious to develop and tedious to follow when generating items. They do, however, prove to be more effective in controlling writers' interpretations and creativity especially for a more complex skill. Therefore, they do produce a more clearly defined behavior domain. (Author/GK)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED206714

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☒ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

An Experimental Comparison of the Effectiveness of Two Domain Definition Strategies for Criterion-Referenced Tests

David M. Logsdon
Assistant Professor
Catholic University

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

D. M. Logsdon

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

TM 810 619

Paper presented at the annual meeting of the Eastern Educational Research Association. Philadelphia, 1981.

Introduction. This research represents an experimental study focused on some key theoretical postulates concerning domain definition strategies of criterion-referenced tests. Popham (1978) defines a criterion-referenced test as an instrument "used to ascertain an individual's status with respect to a well-defined behavior domain" (p. 93). How well the behavior is defined depends on the explicitness of the domain specifications. If the specifications are adequately explicit, then, theoretically, different item writers following the same specifications should generate items which are very similar. As the specifications become less explicit, writers' interpretations become more subjective and items vary across writers. This study empirically investigated the homogeneity or similarity of items based on the same specifications but generated by different writers.

In addition, the study examined the effectiveness of two different domain definition strategies in achieving homogeneity of items. At least seven different strategies have been developed for specifying behavior domains; objectives, amplified objectives (Popham, 1974; Baker, 1974), IOX test specifications (Popham, 1978, 1980), item transformations (Bormuth, 1970; Anderson, 1972; Finn, 1975), item forms (Hively et al., 1973; Osburn, 1968), algorithms (Scandura, 1973, 1977), and mapping sentences (Guttman, 1970, Castro & Jordan, 1977; Jordan, 1978; Schlesinger, 1978; Berk, 1978).

Berk (1980) rated these strategies in terms of "eight practicability factors." He judged the simple objectives and Instructional Objectives Exchange test specifications among the most practical with the latter offering much more clarity and precision to item writers. The IOX strategy,

however, was rated less rigorous and precise than item trasformations, item forms, and algorithms. The present study empirically tested this argument. To what extent do item writers following the IOX domain definition strategy for a cognitive skill generate items that are more homogeneous than items generated from the simple objectives strategy for the same skill?

These questions address the issue of the quantity and quality of specifications needed to achieve clarity of the definition and precision among the items. The IOX strategy contains at least four parts: the general description, sample item, stimulus attributes, and response attributes (Popham, 1978). For a relatively simple skill, say selecting third grade compound words, it requires a substantial amount of time and deliberation to delineate all the specifications assumed necessary to produce homogeneity among writers' items. For a higher level, more complex skill, the effort can become formidable. After all of these specifications are formulated, however, the actual homogeneity of the items remain in question. There is also the possibility that this elaboration is not necessary. Maybe the homogeneity can be achieved with an abreviated definition, i.e., a statement of the objective, a few descriptive comments, and a sample item. In reality, item writers rely heavily on the sample. Perhaps the other specifications are superfluous for a simple skill.

Methods. A set of IOX and abreviated domain specifications were developed by the researcher for each of two third grade language skills, selecting paired connectives (either/or; both/and) and compound words

(backyard, bedroom). The abreviated specifications for each skill were then given to three item writers who were instructed to generate four multiple-choice items for each skill or specifications set. Likewise, the IOX definitions for each skill were given to three item writers with the same instructions. (See attached specifications.) The design is graphically displayed below.

Design

	IOX Definition	Abreviated Definition
Compounds	3 sets of 4 items, n = 12	3 sets of 4 items, n = 12
Paired Connectives	3 sets of 4 items, n = 12	3 sets of 4 items, n = 12

The writers were informed in general terms about the nature of the research and were advised to assume they were writing items for a third grade basic skills test. The writers are School of Education faculty and contract personnel at The Catholic University of America. All are experienced in test construction but not third grade language skills. Writers also were asked to rate and comment on the explicitness of the domain specifications as guidelines for item generation.

Once the items were generated, ratings of similarity were made by three different raters who are experienced with measuring these types of language skills. Each rater compared the similarity of one item set, number 1, with another set, number 2; then, set 1 with set 3; and finally, set 2 with set 3. Items were compared within domain definition strategy, rather than across

strategies. Ratings were made from a 7 point scale and referred to 4 dimensions: similarity of directions; similarity of stimulus sentences; similarity of responses; and overall similarity of the cognitive skill measured. Each of the first 3 dimensions was broken down to variables such as sentence structure, word level, semantics, and format. (See attached rating form.) Analysis was based on means averaged across pair comparisons and raters.

Results. 1. To what extent do item writers following the abreviated or simple objective (plus sample item) strategy generate homogeneous items? Below are the mean similarity ratings averaged across raters for the four dimensions and two skills rated.

Table 1

Mean Similarity Ratings of Item Parts by Pair Comparisons of Item Sets and Skills - Abreviated Specifications

	Compounds				Paired Connectives			
Item Set Comparisons	1/2	1/3	2/3	$\bar{X}$	1/2	1/3	2/3	$\bar{X}$
Directions	3.8[a]	5.9	3.[illegible]	4.4	7.0	5.6	5.6	6.1
Stimulus	5.2	4.0	3.8	4.3	5.5	5.4	5.7	5.5
Response	5.2	4.7	4.3	4.7	4.7	6.2	5.7	5.5
Cog. Skill	4.8	4.3	4.2	4.4	4.5	5.8	4.5	4.9

[a]from a 7 point scale with 7 = "identical"

The items measuring compounds were rated, on the average, only marginally homogeneous, $\bar{X}$ = 4.3 to 4.7, on the 7 point scale with 4 being the midway point between similar and not similar. Response attributes were rated somewhat more homogeneous than the other item parts. Points of dissimilarity were:

Directions - 1 set had ambiguous directions; another required a different behavior for 2 of 4 items.

Stimulus - 1 set used more complex sentence structure; 1 set used longer sentences and higher level words; 1 set varied position of blank space.

Response - different type distractors used in 2 sets; higher level words used.

The paired connective items were rated more homogeneous. Directions were quite similar across writers, $\bar{X} = 6.1$; stimulus sentences and responses were rated 5.5; and the overall similarity of the skill was rated 4.9. Points of dissimilarity were:

Directions - 1 set used higher level words and somewhat different semantics;

Stimulus - somewhat different sentence structure and semantics, "neither/nor", "not only/but also".

Responses - 1 set didn't vary position of correct responses

Compound items were rated less homogeneous because there was more complexity to generating the stimulus sentence and distractors. Writers didn't observe the sample item's patterns and take these as prescriptive guidelines. In some cases writers creatively restructured each item part to their liking.

This creativity emerged from the lack of explicit specifications. Writers rated the abreviated specifications as less than "moderately" explicit. They were "ambiguous" and "unclear" in terms of word level, words used, and "what the domain includes".

2. To what extent do item writers following the amplified IOX strategy generate homogenous items?

Table 2

Mean Similarity Ratings of Item Parts by Pair Comparison of Item Sets and Skills - IOX Specifications

	Compounds				Paired Connectives			
Item Set Comparisons	1/2	1/3	2/3	$\bar{X}$	1/2	1/3	2/3	$\bar{X}$
Directions	1.0[a]	5.3	1.0	2.4	1.0	5.5	1.0	2.5
Stimulus	5.9	5.7	5.8	5.8	5.7	5.5	5.6	5.6
Responses	5.2	5.6	5.0	5.3	6.7	5.7	5.5	6.0
Cog. Skill	5.1	5.3	5.2	5.2	5.7	5.5	5.5	5.6

[a]from a 7 point scale, 7 = "identical"

Compound items generated from the IOX strategy were rated more homogeneous than those from the abreviated definition, with the exception of Directions. Stimulus sentences received high ratings and responses somewhat lower; directions were rated low since one writer neglected to include them. Additional dissimilarities were:

Directions - 1 set was longer and used more complex sentences;

Stimuli - some longer sentences used; blank space position varied;

Responses - higher level words; some noncompound words used; no variation of correct response position.

IOX paired connective items also were rated more homogeneous than those from the abreviated definition, but the differences were not as great as those for compound words. Mean ratings for stimulus sentences, responses, and the overall cognitive skill approached the high end of the scale, between 5.6 and 6.

However, the same writer overlooked directions. Other dissimilarities were:

Directions - 1 writer used 2 sentences rather than 1;

Stimulus - longer, more complex sentences with higher level words;

Responses - no variability of correct responses.

The IOX strategy generated more homogeneous items across writers than the abreviated strategy, but the compound items still did not reveal a high degree of similarity. Paired connective items could have been more homogeneous in terms of stimulus attributes, and Directions were completely omitted from one item set. Writers who followed the IOX definitions found the specifications highly explicit but had problems with some of the terms. One writer claimed they were "too technical."

Discussion. The elaborated IOX specifications are laborious to develop and tedious to follow when generating items. After much deliberation and review in formulating these, writers still pointed out some errors and unclear instructions. However, they do prove more effective in controlling writers' interpretations and creativity, especially for a more complex skill. Hence, they do produce a more clearly defined behavior domain.

This researcher was surprised, however, that more homogeneity was not achieved. Writers apparently need more than written instructions. Two suggestions can be offered. A short training session in which writers are advised of the prescriptive nature of the specifications, with each attribute

being explained. Secondly, a checklist could be supplied enabling each writer to verify that each item conformed to the specifications. Something like:

Directions - same as Sample Item ✓

Stimulus

Attributes - Sentences 10-13 words ✓

Simple structure ✓

3rd grade words ✓

blank space at end ✓

In addition, the specifications should be piloted and evaluated by writers before their applications. The homogeneity of items could be rated, as in this study, to identify needed revisions.

Further research should investigate the influence of writers who are experienced in the type skills being measured. When writers have experience with these skills, more homogeneity could be attained, and they may require fewer specifications. Other skills should be included, especially higher level application skills. The Item-Specification Congruity should be assessed. Implications from this study suggest less than an optimal match with amplified specifications and great variability in the defined domain with abreviated specifications. In addition, the psychometric behavior of the items stands as an important issue. To what extent do items from the different strategies compare in terms of discrimination and reliability? Do items which are not highly homogeneous discriminate poorly and yield low reliabilities?

References

Anderson, R. C. How to construct achievement tests to assess comprehension. Review of Educational Research, 1972, 42, 145-170.

Baker, E. L. Beyond objectives: Domain-referenced tests for evaluation and instructional improvement. Educational Technology, 1974, 14, 10-16.

Berk, R. A. The application of structural facet theory to achievement test construction. Educational Research Quarterly, 1978, 3, 62-72.

Berk, R. A. A comparison of sixcontent domain specification strategies for criterion-referenced tests. Educational Technology, 1980, 20, 49-52.

Bormuth, J. R. On the theory of achievement test items. Chicago: University of Chicago Press, 1970.

Castro, J. G., & Jordan, J. E. Facet theory attitude research. Educational Researcher, 1977, 11, 7-11.

Finn, P. J. A question writing algorithm. Journal of Reading Behavior, 1975, 7, 341-367.

Guttman, L. Integration of test design and analysis. In Proceeding of the 1960 invitational conference on testing problems.

Hively, W., Maxwell, G., Rabehl, G., Sension, E., & Lundin, S. Domain-referenced curriculum evaluation: A technical handbook and a case study from the MINNEMAST Project (CSE Monograph Series in Evaluation, No. 1). Los Angeles: Center for the Study of Evaluation, University of California, 1973.

Jordan, J. E. Facet theory and the study of behavior. In S. Shye (Ed.) Theory construction and data analysis in the behavioral sciences. San Francisco: Jossey-Bass, 1978.

Osburn, H. G. Item sampling for achievement testing. Educational and Psychological Measurement, 1968, 28, 95-104.

Popham, W. J. Selecting objectives and generating test items for objectives-based tests. In C. W. Harris, M. C. Alkin, and W. J. Popham (Eds.), Problems in criterion-referenced measurement (CSE Monograph Series in Evaluation, No. 3). Los Angeles: Center for the Study of Evalluation, University of California, 1974. Pp. 13-25.

Popham, W. J. Criterion-referenced measurement. Englewood Cliffs, NJ: Prentice-Hall, 1978.

Popham, W. J. Domain specification strategies. In R. A. Berk (Ed.), Criterion-referenced measurement: The state of the art. Baltimore, MD: The Johns Hopkins University Press, 1980. Pp. 15-31.

Scandura, J. M. Structural learning I: Theory and research. New York: Gordon and Breach, 1973.

Schlesinger, I. M. Some properties of mapping sentences. In S. Shye (Ed.), Theory construction and data analysis in the behavioral sciences. San Francisco: Jossey-Bass, 1978.

Abreviated Specifications - Paired Connectives

Please construct 4 items, including item Directions, from the following specifications. You may use a dictionary or thesaurus, together with the Word List but no other materials. Do not consult with colleagues. Generate original items.

Domain Specifications

1. Purpose - the items are to be utilized in a Basic Skills-Linguistics test for third graders.

2. Objective - Presented 3 sentences with each featuring paired connectives, the student selects the two that have the same meaning by circling the letters beside the sentences. Two items feature "both/and" connectives and two feature "either/or".

3. Sample Item - Directions: Circle the letters for the sentences that have the same meaning.

   1. a. There are both girls and boys on our team.
      b. There are either girls or boys on our team.
      c. There are girls and boys on our team.

Please construct 4 items, including item Directions, from the following specifications. You may use a dictionary or thesaurus, together with the Word List but no other materials. Do not consult with colleagues. Generate original items.

Domain Specifications

1. Purpose - the items will be utilized in a Basic Skills - Word Meaning test to be administered to third graders.

2. Objective - Presented a simple incomplete sentence which defines an omitted compound word, the learner selects the defined compound word from a list by circling a coded letter. Defined compounds must be legitimate words which appear in standard dictionaries. Component words are at the third grade level.

3. Sample item. Directions: Circle the letter for the best word that finishes the sentence.

   1. A bottle for water is called a _____.

      a. waterbottle

      b. watertop

      c. bottlecap

      d. waterfall

Please construct 4 items, including item Directions, from the following specifications. You may use a dictionary or thesaurus, together with the Word List but no other materials. Do not consult with colleagues. Generate original items.

Domain Specifications

1. Purpose - the items will be utilized in a Basic Skills-Word Meaning test to be administered to third graders.

2. Objective - Presented a simple incomplete sentence which defines an omitted compound word, the learner selects the defined compound word from a list by circling a coded letter.

3 Learning Domain - Intellectual Skill, Defined Concept

4. Stimulus Attributes: brief, simple, incomplete sentences of 7 - 10 words are used. The stimulus defines the compound word by featuring the components of the compound in a highly descriptive definition. The stimuli end with the words "is called a" or "is called the" followed by a blank space. Component words are at the third grade level, and compound words must appear in recent editions of standard dictionaries.

5. Response Attributes - Three plausible distractors and one correct choice are featured in vertical format. Each distractor contains one of the component words of the compound or both of them in reverse order. Two distractors should be "legitimate" words and one may be a fabricated foil. Component words are at the third grade level. Correct choices vary in position across items.

6. Sample item. Directions: Circle the letter for the best word that finishes the sentence.

   1. A bottle for water is called a _____.

      a. waterbottle
      b. watertop
      c. bottlecap
      d. waterfall

Please construct 4 items, including item Directions, from the following specifications. You may use a dictionary or thesaurus, together with the Word List but no other materials. Do not consult with colleagues. Generate original items.

Domain Specifications

1. Purpose - the items are to be utilized in a Basic Skills-Linguistics test for third graders.

2. Objective - Presented 3 sentences with each featuring paired connectives, the student selects the two that have the same meaning by circling the letters beside the sentences.

3. Learning Domain - defined concept.

4. Stimulus Attributes - Each item features 3 choices in vertical format. Each choice represents a simple sentence of 6 - 14 words, a compound sentence of the same length, or 2 simple sentences. Two items feature "both/and" and 2 "either/or". Each of the "both/and" items feature as one choice the form: "There are both girls and boys on our team." The corresponding choices in these items will be of the form: "There are girls on our team and there are boys", or "There are girls on our team. There are also boys." The third choice for these items will be an "either/or" choice about the same topic.

   The "either/or" items should appear in the form: "You may have either fruit or cake for desert." with the corresponding choice in the form: "You may have fruit for desert, or you may have cake." The third choice is a "both/and" choice about the same topic.

   The positions of correct or corresponding choices vary across items. Words should be at the third grade level.

5. Response Attributes - The examinee circles the two letters a, b, or c which indicate sentences or choices that correspond in meaning.

6. Sample item - Directions: Circle the letters for the sentences that have the same meaning.

   1. a. There are both girls and boys on our team.
      b. There are either girls or boys on our team.
      c. There are girls and boys on our team.